Journal of Behavioral and Experimental Economics 72 (2018) 28–39

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Teams in a public goods experiment with punishment☆

Heike Auerswald[a], Carsten Schmidt[b], Marcel Thum[a,*], Gaute Torsvik[c]

[a] Faculty of Business and Economics, TU Dresden, D-01062 Dresden, Germany
[b] Department of Economics, University of Mannheim, D-68131 Mannheim, Germany
[c] Faculty of Social Sciences, University of Oslo, Postboks 1095 Blindern 0317 Oslo, Norway

ARTICLE INFO

JEL classification:
C72
C92
H41

Keywords:
Public good
Group decision-making
Punishment
Experiment

ABSTRACT

Challenges addressed in global politics, such as climate change, maritime piracy and fighting highly contagious diseases, concern global public goods. The related policy decisions are primarily made by teams. In contrast, economic models of global public goods typically assume a single rational decision maker. We use a laboratory experiment to compare team decisions to the decisions of individuals in a finitely repeated public goods game with and without a costly punishment option. Teams of three participants coordinate on decisions either by majority or unanimity rule. We find that, in the absence of a punishment option, teams deciding by the unanimity rule contribute slightly more than individuals. With the punishment option, unanimity teams choose significantly lower levels of punishment and exhibit anti-social punishment less frequently than individuals. A possible explanation is the elimination of extreme preferences for punishment through the coordination process in teams.

1. Introduction

International policy challenges often concern global public goods. Climate change, as one example, is partly attributed to the emission of greenhouse gases. The mitigation of these emissions is a contribution to a global public good. Each country bears the cost of its own mitigation effort, but the benefits accrue to all countries. Another example of an international public good is the fight against maritime piracy. Each country that sends convoying ships to the Horn of Africa makes the passage for all cargo vessels in this area safer. Again, the benefits accrue to traders from all over the world, while the costs are born by the countries providing the convoying ships. For such global public goods, standard economic theory predicts an under-provision due to free-rider behavior in the absence of global regulation. A huge body of experimental literature has shown that private contributions to public goods are indeed below the socially optimal level; however, the contributions are much higher than predicted by orthodox game-theoretic considerations. Explanations for this phenomenon range from confusion (Andreoni, 1995; Houser and Kurzban, 2002) and warm-glow (Palfrey and Prisbrey, 1997) to altruism (Goeree et al., 2002) and conditional cooperation (Fischbacher et al., 2001; Fischbacher and Gächter, 2010).

Public goods experiments usually consider single decision makers, while decisions on global public goods are typically made by teams, as the examples above indicate. The ministry of the environment decides on the national mitigation policy, and the ministry of defense and its experts allocate the resources for fighting maritime piracy. This finding raises the question of whether the results from individual decision-making on public good contributions carry over to team decisions. Do teams show the same behavior regarding the private provision of public goods as individuals do? One may expect that the deliberation in teams fosters strategic considerations and that other-regarding aspects are weakened. With altruism, for instance, individuals may care about other individuals but not about more or less anonymous teams. There is a rapidly growing (experimental) body of literature on team decisions analyzing how teams make decisions and whether their decisions differ from those of individuals (Charness and Sutter, 2012; Kugler et al., 2012). This literature, however, has rarely covered the politically important case of public good provision (Huber et al., 2016), whereas there are several papers on the related prisoner's dilemma (Kagel and McGee, 2016).

Our focus is on teams that jointly provide a single public good. We analyze whether teams contribute significantly more or less than individuals in a standard public goods game with and without a costly punishment option. We extend the standard public goods game with

☆ We thank Tim Cason, Martin Kocher, Robert Lehmann, Wolfgang Luhan, Florian Morath, Vai-Lam Mui, Bernd Suessmuth, Matthias Sutter, Michael Weber, Bengt-Arne Wickström and two anonymous reviewers for helpful comments. Financial support from the German Federal Ministry of Education and Research (Grant no. 01LA1106A) is gratefully acknowledged.
* Corresponding author.
E-mail address: marcel.thum@tu-dresden.de (M. Thum).

https://doi.org/10.1016/j.socec.2017.11.001
Received 9 May 2016; Received in revised form 22 August 2017; Accepted 12 November 2017
Available online 14 November 2017
2214-8043/ © 2017 Elsevier Inc. All rights reserved.

four single decision makers to a setting with four teams who jointly provide the public good. Each team consists of three players who have to coordinate on a team decision. We analyze the contributions to the public good, the use of punishment and the final payoff. For the coordination among team members, we use a structured method of team decision-making in the spirit of Gillet et al. (2009). This method allows team members to make proposals, but it does not allow for communication. The proposals are aggregated to a team decision following either a majority rule or a unanimity rule. The lack of direct communication certainly eliminates an important aspect of real-world team decisions. However, this setting allows us to isolate the effects of the decision-making process itself. Therefore, our analysis is one piece of the puzzle. Another piece is the observation of direct communication among team members, which would allow researchers to study the communication effects.

In our experiment, teams deciding by the unanimity rule contribute more to the public good than individuals. Unanimity teams also choose significantly lower levels of punishment and exhibit anti-social punishment less frequently than individuals. In terms of net profits, unanimity teams performed better than individuals.

In Section 2, we develop our working hypotheses. Section 3 describes the experimental setting in detail. Section 4 discusses the treatment effects. In Section 5, we take a closer look at the punishment stage by analyzing disaggregated data on the individual/team level and by distinguishing between social and anti-social punishment. Section 6 concludes.

2. Literature and hypotheses

Charness and Sutter (2012) and Kugler et al. (2012) provide extensive surveys on team decision experiments. According to Charness and Sutter (2012, p. 171), the majority of research concludes that teams tend to behave more in accordance with game-theoretic predictions than individuals. The authors identify three reasons for the difference between decisions made by teams and individuals. First, individual knowledge is aggregated in teams, and thus, teams make qualitatively better decisions (e.g., investment decisions). Second, teams exhibit more detailed reasoning when making strategic decisions (e.g., in the beauty contest game and the trust game). Teams are better able to anticipate the reaction of the other player and his/her best strategy. Third, teams have a stronger focus on payoffs. Fairness and reciprocity appear to play a minor role in team decision-making.

Kugler et al. (2012) refer also to results from social psychology addressing team decision-making. Most of this literature analyses behavior in a prisoner's dilemma situation and stresses that individuals behave differently as sole decision makers compared to deciding as members of a team. In social psychology, this difference in behavior is usually referred to as the *discontinuity effect*. Two main motives for the discontinuity effect are `greed' and `fear' (Wildschut et al., 2003). 'Greed' refers to a player's stronger focus on payoffs in team decisions. Greed is explained by either the higher anonymity within a team that provides shelter from social punishment or the social support within a team for self-interest behavior (social support of shared self-interest hypothesis; see also Kugler et al. 2007). The second motive, `fear', refers to the expectation of decision makers that teams act more competitively and less cooperatively (schema-based distrust hypothesis). In a prisoner's dilemma situation, decision makers tend to expect teams to act more selfishly and thus to defect more often; they fear being exploited and protect themselves by choosing defection as well. In the public goods game, which is a variant of the prisoner's dilemma problem, the discontinuity effect would suggest that teams contribute less.

Economic research has also investigated team decisions and generally finds that teams behave more in accordance with game-theoretic predictions than individuals. The interactive tasks for which this effect was found include the ultimatum game (Robert and Carnevale, 1997; Bornstein and Yaniv, 1998), the dictator game (Luhan et al., 2009), the beauty contest game[1] (Kocher and Sutter, 2005), the centipede game (Bornstein et al., 2004), the gift-exchange game (Kocher and Sutter, 2007), the trust game (Kugler et al., 2007), the finitely repeated prisoner's dilemma game (Kagel and McGee, 2016), the sequential market game (Stackelberg duopoly, Cardella and Chiu, 2012) and the signaling game with limit pricing and market entrance (Cooper and Kagel, 2005). There are also at least four studies in which teams behave less in accordance with game theory or are less able to process information efficiently. Cason and Mui (1997) were among the first to study experimentally the decisions of teams in an economic framework. In a dictator game, they found: ``*[…] that when a team consists of members who have made different individual offers* [in a previous individual stage of the game, authors' note]*, the team offer tends to be dominated by the more other-regarding member*' (Cason and Mui, 1997, p. 1477). They used an experimental setting with 2-person teams and face-to-face communication. Luhan et al. (2009) repeated the game in a different environment (communication via electronic chat, 3-person teams). They found that teams act more selfishly and that the most selfish player in a team has the strongest influence. The second study that deviates from the main stream of literature is that by Cox and Hayne (2006), who studied a common value auction with risky outcomes. The authors identified a *curse of information*. If additional information is provided on the value of the auctioned item, individuals and teams bid less rationally. In addition, this curse of information effect is stronger for teams. Third, Sutter et al. (2009) show that teams suffer more often from the winner's curse, as teams remain longer in auctions and pay higher prices than individuals. The fourth and most recent study is by Müller and Tan (2013). The researchers established a sequential 2-player market game (Stackelberg duopoly) and found no significant difference between individuals and teams in a 1-shot game. In the repeated game, team decisions were less in accordance with game-theoretic predictions than were individual decisions.

Among the interaction tasks that have been implemented with teams, the finitely repeated prisoner's dilemma game by Kagel and McGee (2016) and the public goods game with an option for ostracism by Huber et al. (2016) are closest to our setting. Kagel and McGee (2016) compared the decisions of 2-person teams playing against each other. Members of a team were allowed to communicate via chat for a certain time span before making a decision. They found that teams cooperated significantly less than individuals in the first super-game, whereas, in the subsequent super-games, the level of cooperation among teams achieved or exceeded the level of individuals. However, as the decision on cooperation in the prisoner's dilemma is binary, there is no information on how strong the willingness to cooperate is. Huber et al. (2016) compare the decisions of 2-person teams who can communicate face-to-face with those of individuals in a standard public goods game. They extend the standard setting by allowing players to vote for the exclusion of other players subsequent to the contribution stage, a specific form of binary punishment option. They find no significant difference in contributions between teams and individuals in the absence of the punishment option and lower average contributions by teams when voting for exclusion was allowed.

We implement a standard public goods game that

(i) extends the binary perspective of a prisoner's dilemma (cooperation or no cooperation) to a scale of cooperation levels (contributions to the public good),
(ii) provides a perfectly controlled decision-making process (coordination via proposals instead of face-to-face communication), allowing us to compare different decision-making rules and
(iii) allows for costly punishment.

[1] Groups tend to perform better during the game due to improved reasoning abilities but not in the first period.

In the contribution stage and the punishment stage, we apply two different decision-making rules: majority and unanimity rule. Both voting procedures have been intensively studied, for instance, in the context of jury decisions. Theoretical and empirical findings in this context show that the two decision-making rules may imply different incentives for strategic voting and consequently lead to different results (e.g., Feddersen and Pesendorfer, 1998, 1999; Guarnaschelli et al., 2000; Bond and Eraslan, 2010; Goeree and Yariv, 2011).[2] These results primarily refer to voting situations where voters have common preferences; therefore, these results are not directly applicable to our scenario, but the strategic interaction with majority and unanimity rules may be similar in our setting.

We address the following research question: Do decisions made by teams differ from those made by individuals in a public goods setting with and without punishment? To answer this question, we examine the level of cooperation (contributions), and we test whether the punishment behavior of teams is closer to standard game theory with rational and selfish preferences.

Our point of reference is the standard economic model of private provision to public goods. In this framework, players maximize their own income, and it is individually rational to not contribute to the public good. Because all players encounter the same incentives, no contributions are made in equilibrium. This result does not change when teams instead of individuals play against each other.

However, in public goods experiments with individual decision makers, positive contributions and a substantial level of cooperation are observed, which indicates that players are partly driven by other factors than their own material payoff, such as altruism, fairness or reciprocity (see Ledyard, 1997). Based on the experimental research on team decisions, we would expect teams to behave in a more competitive and self-oriented manner than individuals (see Charness and Sutter, 2012). The evidence also supports that teams behave more in accordance with game-theoretic predictions in many interactive tasks. Both conclusions from the literature point to lower contributions by teams compared to individuals in a public goods game.[3] Thus, our working hypothesis addressing teams and cooperation is:

H1. Teams contribute less.

Based on the reasoning behind hypothesis H1, we also expect teams to punish less in the presence of a punishment option. Punishment is costly and destroys resources. It also has the characteristics of a public good. If the investment in punishment by one team or individual motivates the punished (and maybe others who observe the punishment) to contribute more, the whole group will profit from the then higher contributions. Thus, the costs of punishment are private; however, the benefits are public, and we expect a free riding incentive to be at work. Our second working hypothesis is:

H2. For a given level of contributions, teams punish less.

While there is an incentive to free-ride on the others' contributions to a public good, contributions increase *welfare* from a collective perspective. In accordance with recent approaches (Gächter et al., 2008; Ambrus and Greiner, 2012), we also compare the welfare of different treatments. For our welfare analysis, we abstract from potential warm glow effects and altruism, which may enter the welfare function, and focus on payoffs as our welfare measure. The overall outcome depends on both contributions and punishment. Therefore, a corollary of hypothesis H1 is:

C1. Teams yield lower payoffs in the absence of punishment.

In the absence of punishment, payoffs are a linear function of aggregate contributions. Thus, a corollary of hypothesis H1 is that average payoffs are lower in the team treatments. In the presence of punishment, the overall outcome is ambiguous. Payoffs depend positively on contributions but are diminished by costly punishment. Because the two effects are countervailing, and because punishment may depend on contributions, it is far from obvious whether teams will generate higher or lower aggregated payoffs than will individuals.

3. Experimental design and procedure

The general framework is a standard public goods game with and without punishment (Ledyard, 1997; Fehr and Gächter, 2000, 2002). An experimental group consisted of four *individuals* or four *teams*. Each team consisted of three players. The set of four teams that interacts in the provision of the public good is referred to as a *group*. Each individual/team received the same endowment (20 tokens) at the beginning of each period and was required to decide how many (integer) tokens to invest in the public good. The marginal per capita return for each token invested in the public good was 0.4, which accrued to each of the four individuals/teams.[4] All tokens not invested in the public good were retained by the individual/team; this was equivalent to a marginal return of 1 for the investment in the private good.

In the public goods game with punishment, the contribution stage was followed by a punishment stage. In accordance with the literature on punishment, we implemented the following punishment mechanism: a 3-token reduction in the payoff of another player costs the punishing player 1 token (see Gächter et al., 2008).

We established three between-subject treatments: an individual treatment (IND), a majority team treatment (MAJ) and a unanimity team treatment (UNA). In the majority team treatment, at least two of three team members were required to agree on the levels of contribution and punishment to form a team decision. In the unanimity teams, a team decision required the agreement of all members on contribution and punishment.

The experiment consisted of three parts (see Fig. 1). In *Part I*, all participants played an individual 1-shot public goods game (period 0). This 1-shot game reveals information about the type of player, e.g., her preferences for contributions. No feedback was provided in Part I, to avoid learning at this stage. At the beginning of *Part II*, participants were randomly assigned to the treatments (IND, MAJ and UNA) and to the teams for the UNA and MAJ treatments and the instructions for this part were distributed. Participants played a 10-period standard public goods game as a team member or an individual player with fixed matching (periods 1–10). After each period, participants received feedback concerning their results and anonymized information on the decisions of the other individuals/teams. *Part III* was implemented as a surprise restart.[5] Participants remained in their treatment, and teams were re-matched within their matching group of 4×3 players such that every participant was assigned to new teammates. This procedure was common knowledge. The re-matching ensured that participants had no prior information concerning the behavior of their teammates. In Part III, participants played a 10-period public goods game with fixed matching and costly punishment (periods 11–20). After each contribution and each punishment stage, participants received feedback on their results and anonymized information on the decisions of the other individuals/teams. After Part III had been completed, we disclosed the

[2] Feddersen and Pesendorfer (1998) show, for instance, that due to strategic voting, a jury deciding unanimously convicts an innocent defendant with a higher probability in comparison to a jury deciding by majority. In addition, Bond and Eraslan (2010) note that this result may crucially depend on the exogeneity of the issue that is put to vote. Goeree and Yariv (2011) find that collective decisions differ with different decision-making rules when communication is restricted (no deliberation allowed).

[3] In the textbook model for the private provision of public goods, the punishment option makes no difference as the game is finitely repeated and investment in punishment is also a private provision to a public good. Hence, the hypothesis about contributions should be independent of the punishment option.

[4] The individual payoff of a team member is equal to the team result. Technically, the experimenter tripled the team's profit and awarded an equal share to each team member. Thus, the incentives in the individual and team treatments were identical.

[5] Subjects were allowed to quit after Part II (no one did so). Part III payoffs were in addition to the previous parts.

Treatment	Part I Contribution	Part II Contribution	Part III Contribution & Punishment (Stage 1) — (Stage 2)
MAJ	Individual decision-making	Majority decision-making in teams	
UNA		Unanimous decision-making in teams	
IND		Individual decision-making	
Periods	0	1-10	11-20
Feedback*	no	yes	yes

Fig. 1. Structure and sequence of the experiment. Note: * Feedback after each proposal and decision.

results of Part I.

We implemented a structured method of team decision-making in the spirit of Gillet et al. (2009). At the beginning of each period, each team member could propose a contribution to his/her teammates. Specifically, each team member was required to type in a number (0⋯20) as his/her proposed contribution to the public good. Once all three team members had entered their numbers, the proposals were shown to the team. Under majority rule, two of the three proposals were required to be equal to form a decision. Under unanimity rule, all three proposals were required to be equal. Except for making proposals to the other team members, we did not allow for any communication between team members. Admittedly, face-to-face communication is an important part of real-world team decisions. However, this minimal structured type of communication in the spirit of pure coordination allowed us to isolate the effects of the decision-making process (Gillet et al., 2009).

If an agreement was achieved according to the team decision rule, it was automatically implemented as the team decision. If no agreement was achieved, a new round of proposals started. After a maximum of 10 rounds of proposals without reaching a decision, a default rule applied. Then, the decisions of another team that had achieved a compromise was randomly selected as the undecided team's choice, and the team's payoff for this period was set to zero. Participants had full information about the default rule. The constraint that there was a maximum of 10 rounds of proposals was never binding in the majority treatment; it became binding in less than 2% of the cases in the unanimity treatment. In Part II and stage 1 of Part III (see Fig. 1), coordination failures occurred in only 15 cases of a total of 800 unanimity team decisions.[6]

In Part III, players (individuals/teams) could penalize other players at a cost. When all decisions regarding contributions had been made in a period, each participant received anonymous feedback concerning the contributions of the others. Then, players (individual/teams) could decide about the allocation of punishment points to the other three players. Reducing the payoff of another player by three tokens cost the punishing player 1 token. In team decisions, each member of a team could propose an investment in punishment, for example, team A: 0 tokens, team B: 1 token and team C: 3 tokens. If the punishment schedules of two (three) team members were identical, they formed the team decision. This process was automatically implemented by the program according to majority (unanimity) rule. As in the contribution stage, the default rule was designed in a manner to minimize strategic behavior of team members. If a team had not found agreement after 10 rounds of proposals, the software randomly selected one of the decisions of the other teams in the group. The team had to bear the costs of this random decision. Teams failed to agree on the assignment of punishment points in 8 of 400 cases of unanimous decisions. In the majority treatment all teams did reach an agreement in all cases.[7]

The experiment was programmed and conducted using z-tree software (Fischbacher, 2007). Sessions were held at the experimental laboratory at the University of Mannheim from September to November 2012. Subjects were recruited using a database of volunteers (Greiner 2003, 2015). In all treatments, subjects were invited to participate for 90 minutes. In general, individual treatment sessions lasted approximately 60 minutes; sessions with the majority treatment lasted 75 minutes, and those with the unanimity treatment lasted a maximum of 90 minutes. In total, 280 students, approximately half of which were bachelor-level business administration students, participated in the experiment. After being randomly seated in separate cubicles in front of a computer, which was subsequently used to run the experiment, participants received the instructions for Part I. The experiment began after all participants had passed a quiz ensuring they all had understood the instructions. The instructor answered questions individually and in private. No communication among participants was allowed during the experiment. The exchange rate for tokens was 2 Eurocent per token. Payments were made anonymously in cash immediately following the experiment, with an average payoff of 11 Euro.

4. Results

Table 1 contains the summary statistics on the group level for average contributions, average punishment (when applicable) and average net profits in each treatment over all periods. In total, average contributions are higher, and average punishment is lower in the team treatments than in the individual treatment. This finding results in higher net profits in the team treatments as well. Comparing Parts II and III, the average contributions are more than 1.5-times higher when a punishment option exists.

Coordination among team members was achieved fairly rapidly. Under majority rule, the teams required, on average, 1.6 and 1.3 rounds to reach a decision without and with punishment, respectively. In the much more complicated treatment with unanimity, 80% of teams reached agreement in 5 rounds or less. Here, the average number of rounds was 3.9 and 2.7 without or with punishment, respectively. As stated earlier, we observed coordination failures only in the unanimity

[6] The default rule is a compromise between manageability and avoiding feedback effects from the default rule on the team decisions. In the majority treatment, a single player can never cause a default. In the unanimity treatment, however, a default implements non-extreme contributions for the defaulting team on average. If there are two players with extreme preferences in a team and one with preferences near the average, the third player could use the default option to achieve a contribution closer to his own preferences. We observe this behavior in 6 of 15 cases. Most defaults are simply due to coordination failures.

[7] Fixing a certain level of punishment, e.g., `no punishment', as a default would distort the team decision. This restriction would provide the player who prefers the default outcome an incentive to ensure that the team finds no agreement.

Table 1
Summary statistics: average contributions, punishment and net profits.

		Part I 1-shot PG w/o feedback		Part II repeated PG with feedback		Part III repeated PG with punishment and feedback	
	N (1)	Avg. contributions (2)	Avg. contributions (3)	Avg. net profits (4)	Avg. contributions (5)	Avg. punishment[a] (6)	Avg. net profits (7)
UNA	10	9.075	9.775	25.235[b]	16.240	0.533	27.272
		(4.116)	(5.140)	(3.976)	(5.748)	(0.763)	(5.630)
MAJ	10	9.617	8.978	25.387	14.950	0.588	26.620
		(3.315)	(5.893)	(3.536)	(5.277)	(0.749)	(4.716)
IND	10	8.950	7.563	24.538	14.890	1.108	24.511
		(3.551)	(5.314)	(3.188)	(6.170)	(1.392)	(7.387)
Periods		0	1–10	1–10	11–20	11–20	11–20

Note: standard deviations in parentheses, N: number of groups, UNA: unanimity treatment, MAJ: majority treatment, IND: individual treatment.
[a] Average number of assigned punishment points, equaling the average cost of punishment.
[b] Due to a small number of default cases in the UNA treatment, the average net profits of teams in the UNA treatment are lower than those in the MAJ treatment, although contributions, on average, are higher.

team treatment, and there, in less than 2% of the decisions. One reason for the high coordination is certainly that there are focal points, despite the broad range of possible contributions from 0 to 20. Most participants suggested a contribution of 0, 10 or 20.

We also test whether the random assignment to the three treatments subsequent to Part I was successful. A 2-sample t test shows no significant differences between treatments; we can refrain from a possible matching bias.[8]

We now test the hypotheses from Section 2. The subsequent analyses are either implemented at the individual/team level or at the group level.

H1. Teams contribute less.

Fig. 2 depicts the average contribution for each round for the two team treatments (MAJ, UNA) and the individual treatment (IND). The left panel refers to Part II without a punishment option and the right panel to average contributions in Part III with punishment. The two panels exhibit well-known properties in a finite horizon repeated public goods game. In the absence of a punishment option, cooperation erodes over time, and contributions decline. With a punishment option, cooperation sustains (Ledyard, 1997; Fehr and Gächter, 2000, 2002). An initial inspection of the curves for the three treatments suggests that teams contribute more in nearly all periods in Part II. Moreover, for Part II, the two team treatments are relatively close together in the beginning and split after period 5 with UNA lying above MAJ. In Part III, the UNA teams contributed more than the MAJ teams and the individuals, the two being very close together throughout all periods.

We will now proceed in two steps. First, we use a cross-section analysis to investigate the effects on the individual/team level for the first period of part II (period 1) and part III (period 11), respectively. Second, we implement a panel regression to analyze the effects at the group level, including data from all periods.

The cross-section analysis provides a general idea of the effect of 'being part of a (majority or unanimity) team' on the contributions by analyzing the average difference between the two team treatments and the individual treatment. To eliminate any group-specific or learning biases, we only use the data of the first period of each part of the experiment. At this point, teams and individuals did not yet have information on the preferences of the other teams or individuals in their group and had no specific experience with the setting.[9] We implement a standard OLS regression with robust standard errors (see Appendix B for details). The results for the cross-section are reported in Table 2 (Columns 1 and 3). We explain the average contribution at the group level with a dummy for each team treatment and the contributions made in the 1-shot game in Part I of the experiment (period 0). The contribution levels in Part II and Part III are strongly correlated with the contributions in the 1-shot game.[10] Assuming that the individual decisions from Part I are a proxy for a participant's selfishness or other-regarding preferences, this result suggests that initial preferences are an important driver behind contribution decisions. The size of the effect is large as well. A 1-token increase in the contribution in the 1-shot game translates into a 0.6 (0.5)-token greater contribution in the first period of Part II (III).[11]

In contrast to our expectations formulated in hypothesis H1, we find that contributions of unanimity teams in Part II are higher than those of individuals at a significance level of 5.9%. On average, UNA teams contribute 34% more than individuals in the first period.[12] For Part III, we find no statistically significant effect of the UNA- or MAJ-dummy on contributions, which is in accordance with the first impressions from Fig. 2 and the small relative differences between average contributions from the summary statistics.

Table 3 reports the results of the corresponding panel regression over all periods. We use a random effects model because our variables of interest, the UNA and MAJ-dummies and the contribution in the 1-shot game, are time invariant. We also include period dummies and clustering at the group level to control for learning and group-specific effects (see Appendix B for details). Columns 1 and 3 in Table 3 support the results from the first-period cross-section analysis. Contributions are mainly explained by the initial preferences that are revealed in the 1-shot game in Part I (period 0). Here, a 1-token increase in the contributions in Part I results in an increase of 0.8-token in Part II. In addition, teams deciding by the unanimity rule contribute significantly more than individuals in Part II (p-value 0.023). Both correlations are (nearly) undetectable in Part III, in the presence of a punishment option.

H2. For a given level of contributions, teams punish less.

We should observe less punishment for a given level of contributions in a team setting if teams behave closer to standard game-theoretic predictions. Punishment is costly for the punisher and the

[8] See Appendix A for details.
[9] In Part III, all participants had experience with the decision-making process in their treatment but not with decisions regarding punishment.
[10] In the case of teams, we calculate the average 1-shot contribution across the team.
[11] Inspecting the first period *first proposals* on the subject level, we find that a 1-token increase in the contribution in the 1-shot game translates into a 0.5 (0.4)-token higher first proposal in the first period of Part II (III). This effect is highly significant for both parts.
[12] The coefficient of the UNA dummy is 2.598. The average contribution in the individual treatment without punishment equals 7.563 (see Column 3 in Table 1). This result yields a team effect of 34% ($\approx 2.598/7.563 \times 100$).

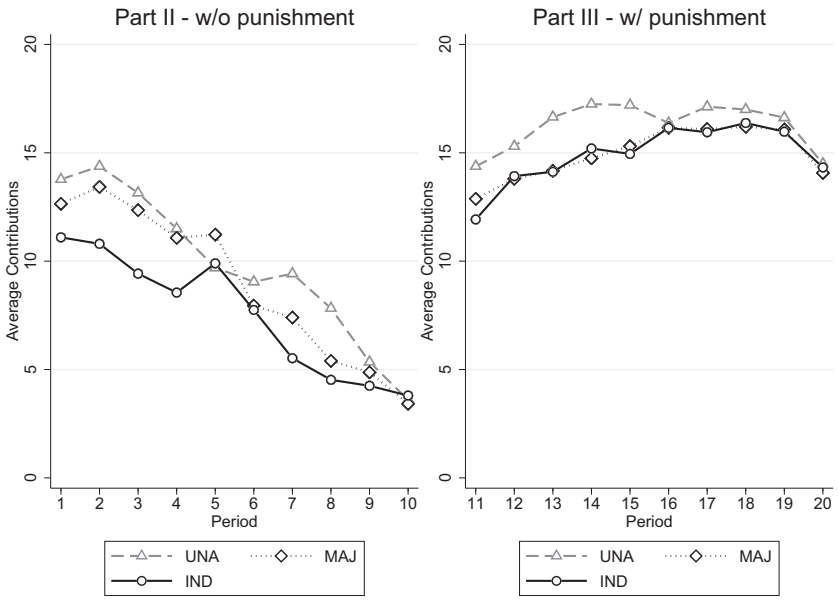

Fig. 2. Average contributions in the treatments with and without punishment.

Table 2
Regression results – first period (individual/team level).

	Part II PG with feedback		Part III PG with punishment and feedback		
	Contributions[a] (1)	Net profits[a] (2)	Contributions[a] (3)	Received punishment[b] (4)	Net profits[a,c] (6)
UNA	2.598^{+} (1.362)	3.160** (1.073)	2.393 (1.443)	−0.524 (0.695)	1.881 (1.605)
MAJ	1.139 (1.327)	1.831 (1.117)	0.645 (1.550)	−0.781 (0.801)	1.191 (1.348)
Contributions		−0.581*** (0.067)		−0.323*** (0.051)	−0.127 (0.083)
Contribution in 1-shot game[d]	0.617*** (0.102)		0.457*** (0.105)		
Intercept	5.576*** (1.449)	33.112*** (1.221)	7.833*** (1.440)	3.281*** (0.598)	24.770*** (1.148)
Observations	120	120	120	120	120
Period Considered	1	1	11	11	11
R^2	0.287	0.412	0.151	0.193[e]	0.030

$^{+}$ p < .10, * p < .05, ** p < .01, *** p < .001,
[a] Standard OLS regression with robust standard errors in parentheses,
[b] Tobit regression (marginal effects) with robust standard errors in parentheses,
[c] Profits net of investment in punishment and punishment received,
[d] Team average for UNA and MAJ treatments,
[e] Pseudo R^2.

punished. If punishment drives increases in contributions in subsequent rounds, this costly punishment is a private provision of a public good to the whole group. Fig. 3 depicts the average investment in punishment[13] for the two team treatments and the individual treatment for each period. The levels of punishment for both team treatments lie beneath the individual level and decline slightly, whereas the level in the individual treatment remains nearly constant over time.

We explore the difference in punishment using a Tobit analysis, because the data are left-censored and, in a majority of cases, individuals/teams choose not to invest in punishment.[14] Again, we proceed in two steps. The cross-section analysis at the individual/team-level data of period 11 in Part III reveals how contributions and the team treatments correlate with the punishment an individual/team received. The results in Table 2 Column 4 show a strong statistically significant negative relationship between the contributions of an individual/team and the punishment received. Thus, higher contributions result in lower punishment. However, we find no significant team treatment effects. This result is not surprising, as in the first period, individuals did not punish in 24 of 40 cases, and teams did not punish in 51 of 80 cases. Thus, we only have very few observations of punishment with nearly no variation.

We obtain the same results at the group level using a random effects Tobit model including all periods of Part III. The results are provided in Table 3 Column 3. Contributions have a negative and strong statistically significant effect on the average investment in punishment. If contributions to the public good increase in a group, the investment in punishment decreases. Again, there is no statistically significant team treatment effect on the level of punishment, which may be a result of the low number of punishment cases in our data. In Section 5, we will decompose the punishment data to derive more insights.

C1. Teams yield lower payoffs in the absence of punishment.

Next, we compare the net profits realized in the different treatments. The net profits of each individual/team are calculated by adding the returns from the public good to the retained tokens and then subtracting the costs of punishing and being punished. Therefore, the following analysis will focus solely on the individual/team level. At the group level, net profits are a linear function of total group contributions and total punishment.

Fig. 4 displays the average net profits over the 10 periods by

[13] The term "investment in punishment" in our setting always refers to the expenditures of the punisher (the assigned punishment points), not to the number of tokens taken from the punished. The latter can be easily calculated by tripling the investment.
[14] In 866 of 1,200 cases, individuals or teams choose not to punish.

Table 3
Regression Results – All Periods (group level).

	Part II repeated PG with Feedback	Part III repeated PG with Punishment and Feedback	
	Avg. contributions[a]	Avg. contributions[a]	Avg. punishment[b]
	(1)	(2)	(3)
UNA	2.111*	1.267	−0.754
	(0.927)	(2.324)	(0.499)
MAJ	0.876	−0.382	−0.422
	(1.225)	(2.005)	(0.490)
Contributions			−0.125***
			(0.027)
Contribution in 1-shot game[c]	0.808***	0.663+	
	(0.128)	(0.376)	
Intercept	4.063**	6.654	2.596***
	(1.425)	(4.207)	(0.523)
Period dummies	yes	yes	yes
Observations	300	300	300
Periods considered	1–10	11–20	11–20
R^2	0.527	0.130	

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.
[a] Random effects model with clustered standard errors in parentheses
[b] Random effects Tobit model
[c] Team average for UNA and MAJ treatments.

treatment with and without punishment. In the treatment without punishment, the net profit curves of the two team treatments lie slightly above that of the individual treatment except for the last period. Because net profits are a linear function of aggregated contributions for Part II, the pattern matches the one of the contributions.

In Part III, the contribution levels do not differ much between the three treatments. Hence, it is primarily the slightly lower level of punishment imposed by UNA teams that leads to higher net profits in the UNA treatment. If we interpret the results in terms of economic welfare, teams with unanimity rule on average generate higher welfare.

Again, we complement the investigation of net profits with a cross-section analysis focusing on period 1. Our independent variables are the contributions of teams or individuals and the two team treatment dummies. Columns 2 and 6 in Table 2 report the results. As expected, higher contributions of a team or individual lead to lower net profits in Part II. This result mirrors the free riding property in our public goods game. We also find a strong statistically significant correlation between the UNA treatment and the net profits in Part II. Hence, the higher contribution levels of UNA teams yield significantly higher net profits in this treatment than in the individual treatment. Again, in Part III, we find no statistically significant treatment effects on net profits.

5. Decomposition of punishment

In the previous section, we have shown that the average punishment is lower in team treatments, although the effect is not statistically significant. Now, we analyze the punishment data in more detail.

We distinguish between the first proposals and the final decisions on punishment in the team treatments to gain more insight into the team decision-making process. Table 4 shows the results for the first proposals in the upper part and, for the final decisions, in the lower part of the table. For both team treatments, the first proposals are, on average, higher than the final decisions on punishment (Column 2). This result indicates that the decision-making process in teams dampens the extreme preferences in favor of punishment.

Lower investment in punishment by teams can arise from two sources. Teams may use the punishment option less frequently, or they may choose lower levels of investment in punishment. To distinguish between the two effects, we split the investment in punishment into (1) the decision to punish at all (probability of punishment) and (2) the investment that is chosen (level of punishment). First, we have calculated the proportion of punishment incidences over the 10 periods for each individual and each team (Table 4, Column 3, lower panel). The probability that an individual exercised the punishment option is 0.345. For teams, this probability is only 0.228 (UNA) or 0.263 (MAJ). This difference, however, is not statistically significant, as the p-values at the bottom of Table 4 indicate. For the proposals (upper panel), the probability of punishment is lower only for the proposals in the UNA treatment.

In addition, we have computed the level of punishment (Table 4, Column 4). This computation is the average number of tokens spent on punishment conditional on having decided to punish at all. Here, the difference between individuals and the two team treatments is statistically significant for the final decisions, although only at a very low level (lower panel of the table). Whenever individuals used the punishment option, they spent approximately 3 tokens, on average, on punishment, while teams only spent slightly more than 2 tokens.

When we compare the relative size of the effects for the probability and the level of punishment between the team treatments and the individual treatment, the two components nearly equally contribute to

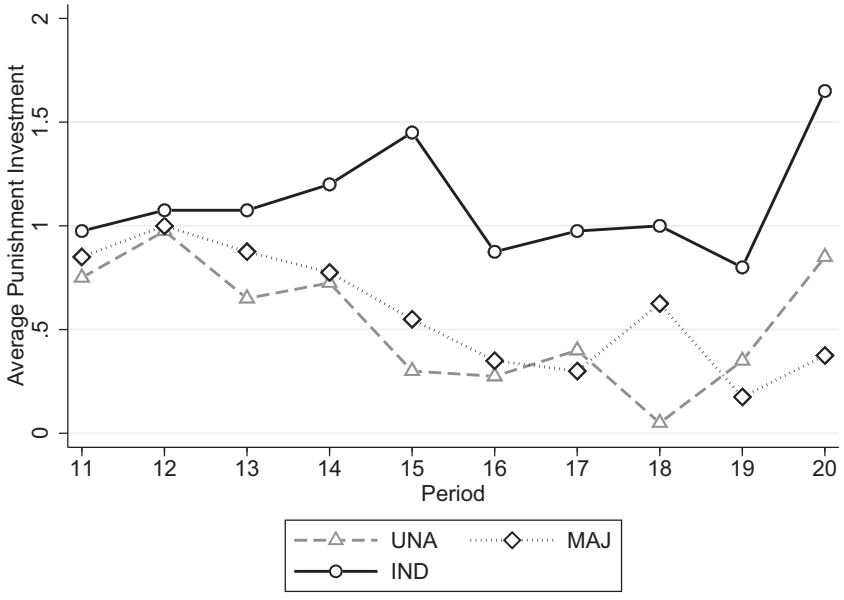

Fig. 3. Average punishment investment.

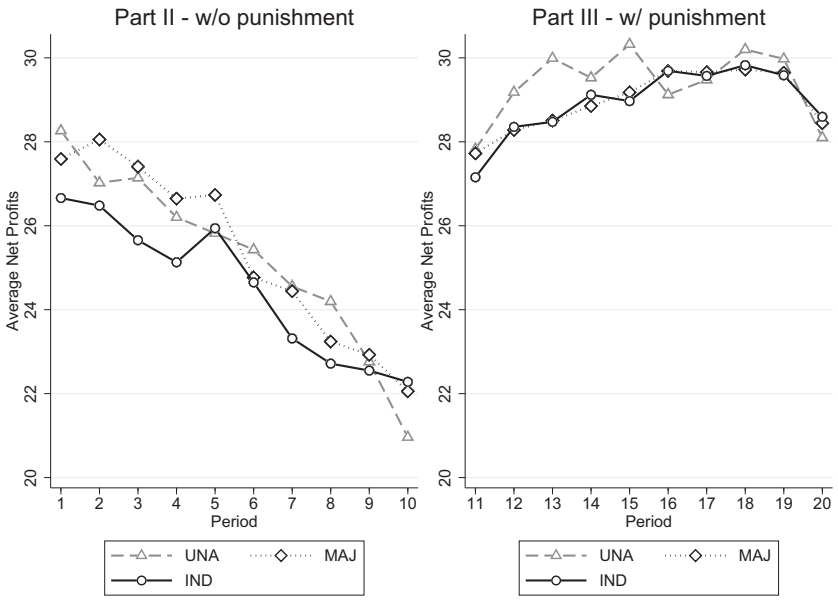

Fig. 4. Average net profits.

Table 4
Decomposition of punishment.

	N (1)	Avg. investment in punishment (2)	Probability of punishment (3)	Level of punishment (4)
UNA- 1st proposals	10	0.609 (0.429)	0.223 (0.190)	2.948 (0.785)
MAJ- 1st proposals	10	0.763 (0.236)	0.310 (0.109)	2.579 (0.676)
IND	10	1.108 (1.105)	0.345 (0.300)	3.174 (1.154)
Mann–Whitney-*U* test[a]				
UNA vs. IND		$p = .450$	$p = .364$	$p = .807$
MAJ vs. IND		$p = .762$	$p = .450$	$p = .241$
UNA- final	10	0.533 (0.458)	0.228 (0.210)	2.350 (0.858)
MAJ- final.	10	0.588 (0.288)	0.263 (0.128)	2.270 (0.694)
IND	10	1.108 (1.105)	0.345 (0.300)	3.174 (1.154)
Mann–Whitney-*U* test[a]				
UNA vs. IND		$p = .273$	$p = .325$	$p = .086$
MAJ vs. IND		$p = .677$	$p = .970$	$p = .064$

Note: standard deviation in parentheses; due to a different level of aggregation, the standard deviations reported in Column 2 differ slightly from those reported in Table 1, Column 6.

[a] Also known as the Wilcoxon rank-sum test.

the difference that is observed.[15] Hence, the difference between the average punishment in the team treatments and in the individual treatment equally stems from a difference in the probability of punishing other players and the difference in the level of punishment that is chosen. From the difference in proposals to the final decisions, we can also conclude that it is not only the mere assortment into a team (`team identity') but also the decision-making process within the team (coordination process) that leads to lower investments in punishment. This result also explains why we find stronger effects in the UNA treatment, where the coordination of all 3 teammates is necessary.

A closer examination of the data also reveals a number of cases of anti-social punishment. It is common in the individual treatment, but it barely ever shows in the team treatments. Anti-social punishment is defined as ``*the sanctioning of people who behave prosocially*'' (Herrmann et al., 2008, p. 1362). In our case, anti-social punishment occurs when a player (individual or team) punishes another player who contributed more. Fig. 5 splits the punishment investment from Fig. 3. The left panel shows the average investment in punishment for each period in all cases where the punished individual or team had contributed less than the punisher (social punishment). Individuals, as well as the UNA and MAJ teams, use punishment in the first periods to enforce cooperative behavior. The use of the option declines over time and does so more strongly in the team treatments. The right panel displays the spending on anti-social punishment. This spending includes all cases where the punisher had contributed less than the punished. The panel exhibits three striking features. First, anti-social punishment is much lower than social punishment, as it occurs more rarely. Second, anti-social punishment increases over time in the individual treatment. The nearly constant average punishment by individuals depicted in Fig. 3 masks that there is an increase in anti-social punishment and a parallel decrease in social punishment. Third and most importantly for our investigation, there is barely any anti-social punishment in the two team treatments. Decision-making in teams appears to eliminate the destructive anti-social behavior.

In Table 5, we repeat the exercise from Table 4 but focus on the cases of anti-social punishment only. Again, the table shows the first proposals in the upper part and the final decisions in the lower part. Column 2 displays the average anti-social punishment and Columns 3 and 4 report the decomposed anti-social punishment (probabilities and levels). Comparing the proposals to the final decisions of teams, we observe the familiar picture from Table 4. The coordination process within teams may lead to lower probabilities and levels of anti-social punishment. For instance, the level of punishment changes from approximately 3 (2.5) tokens in the first proposals to 1.5 (1.6) tokens in the final decisions in the UNA (MAJ) treatment. However, only the difference in the average anti-social punishment between the UNA treatment and the IND treatment shows a weak statistically significant correlation (Column 2).

A possible channel through which punishment may be reduced in team treatments is that the team decision-making process prevents extreme preferences from being translated into final decisions. A first indicator is the higher probabilities and levels of punishment in the first proposals reported in the upper panels of Tables 4 and 5 compared to the final decisions of teams. If the team's median voter prevails and if the median's willingness to punish is below the average willingness to punish and thus the distribution is skewed, teams should exhibit lower punishment levels than the average individual decision. Suppose that the distribution of true preferences over investment in punishment

[15] For example, comparing IND and UNA yields a relation of 0.345/0.228 = 1.5 for the probability and 3.174/2.350 = 1.4 for the level of punishment.

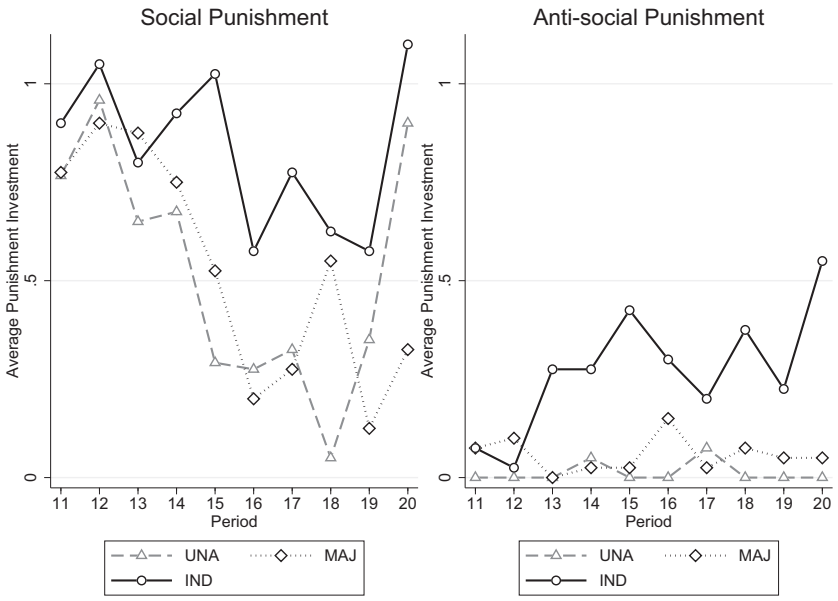

Fig. 5. Social and anti-social punishment.

Table 5
Decomposition of anti-social investment in punishment.

	N	Avg. investment in anti-social punishment	Probability of anti-social punishment	Level of anti-social punishment
	(1)	(2)	(3)	(4)
UNA- 1st proposals	10	0.131 (0.118)	0.047 (0.050)	3.145 (0.778)
MAJ- 1st proposals	10	0.245 (0.162)	0.093 (0.044)	2.536 (1.186)
IND	10	0.273 (0.395)	0.078 (0.110)	4.209 (4.915)
Mann–Whitney-*U* test[a]				
UNA vs. IND		$p = .908$	$p = .908$	$p = .565$
MAJ vs. IND		$p = .305$	$p = .138$	$p = .874$
UNA- final	10	0.013 (0.032)	0.008 (0.018)	1.500 (0.707)
MAJ- final	10	0.058 (0.094)	0.033 (0.049)	1.633 (0.606)
IND	10	0.273 (0.395)	0.078 (0.110)	4.209 (4.915)
Mann–Whitney-*U* test[a]				
UNA vs. IND		$p = .016$	$p = .025$	$p = .295$
MAJ vs. IND		$p = .156$	$p = .289$	$p = .185$

Note: standard deviation in parentheses.
[a] Also known as the Wilcoxon rank-sum test.

within a team is (0, 1, 5). Then, given that all team members strictly reveal their preferences, the median voter theorem predicts that an investment of 1 will be chosen in a simple voting procedure. If these three team members had participated in an equivalent individual treatment, we would have observed an average punishment of 2. Thus, with a skewed distribution, decision-making in majority teams may result in lower punishment. To test whether such skewness could be responsible for our result, we consider the team players' first proposals for punishment in the first round, as these most likely reflect their true preferences. Clearly, the first proposals within a team are not truly independent; however, because we focus on the first period, this is as close as we can be to a true statement of individual preferences. Furthermore, because we do not have a clear theoretic framework to predict the outcome in unanimity teams, we simply extend the median voter assumption to the UNA treatment as well.[16]

Table 6 shows the average difference between the maximum/minimum proposal and the median proposal within a team for the two team treatments. The distribution is skewed. In both treatments, the median proposal is much closer to the minimum than to the maximum proposal, and the difference is statistically significant. This result suggests that teams show a higher degree of economic rationality by dampening the extreme high preferences for punishment.

Table 6
Skewness of preferences: initial punishment proposals to team members in the first period.

	Number of teams	Initial punishment proposals	
		MAJ	UNA
	(1)	(2)	(3)
Maximum–Median	120	0.517	0.692
Median–Minimum	120	0.225	0.208
Difference		0.292	0.483
Wilcoxon matched pairs signed-rank test		$p = .000$	$p = .000$

6. Conclusions and outlook

We analyze team decisions in a public goods setting. The outcomes of our experimental study are mixed and confirm some but not all of the results of previous studies on team decisions in other games. Our results diverge in that teams in general do not behave in a manner more similar to game-theoretic predictions.

In contrast to our initial expectation and our working hypothesis, we find a positive link between the unanimity team treatment and contributions, as well as net profits, in Part II of our experiment (repeated public goods game without punishment). Contributions are also strongly determined by initial preferences (measured by the individual contribution decisions in the 1-shot game in Part I). Interestingly, we find differences in punishment behavior between individuals and teams when we more closely examine the data. Teams tend to choose lower levels of punishment than individuals. Additionally, the punishment of players who contributed more (anti-social punishment) occurs significantly less often in the unanimity team treatment than in the individual treatment. Assuming that punishment is a tool used to foster cooperation, it appears that the coordination process within teams tends to eliminate destructive behavior and leads to better decisions. We also find that, in the team treatments, extreme preferences in favor

[16] A simple OLS regression confirms for both team treatments that the final decisions are mainly driven by the median proposal in the first round. Details are provided in Appendix C, Table C.1.

of punishment are absorbed.

The lower punishment in team treatments may also occur because those who punish are afraid of harming an ``innocent bystander" (a player who voted for higher contributions but was overruled by the teammates). In our setting, it is only possible to punish the team as a whole, and the punisher cannot identify the team members responsible for the low contribution level. To isolate the ``innocent bystander" effect, a mixed setting with treatments where individuals are also playing against teams is needed. Such a setting would allow the exploration of whether the lower punishment arises from teams deciding to punish less or teams being punished less.

Another fruitful avenue for future research could be studying team communication. For instance, the higher contributions in the unanimity treatment could be driven by the coordination efforts, when each team member has veto power. A more elaborated setting, including direct communication, e.g., via chat, would allow the testing of further hypotheses regarding the coordination process in teams.

Appendix A

We check whether the random assignment procedure that sorted the subjects into the treatments after Part I was successful. We compare the individual contributions in Part I over the three treatments using a 2-sample t test with unequal variances.[17] Table A.1 provides the results. None of the three possible pairings of treatments shows any statistically significant difference in the Part I contributions. Hence, it appears justified to assume that the matching procedure was successful.

Table A.1
Comparison of contribution levels in part I (group level) – t test on differences between treatments.

	Part I contributions
IND vs. MAJ	.609
IND vs. UNA	.927
MAJ vs. UNA	.577

Note: p-values based on 2-sample t test with unequal variances.

Appendix B

Cross-section regression model

OLS at the individual/team level with robust standard errors, first period only

$$\overline{x}_i = \alpha_0 + \alpha_1 UNA_i + \alpha_2 MAJ_i + \alpha_3 \overline{x}_{i,I} + \varepsilon_i \tag{B1}$$

$\overline{x}_i$ average contribution by individual/team i
UNA_i, MAJ_i dummies for team treatments (0 for IND, 1 for UNA/MAJ)
$\overline{x}_{i,I}$ average contribution by individual/team i in Part I (individual 1-shot game)
ε_i error term
i index for players (individual/team)

Panel regression model

Random effects at the group level with period dummies and clustered standard errors

$$\overline{x}_j = \beta_0 + \beta_1 UNA_j + \beta_2 MAJ_j + \beta_3 \overline{x}_{j,I} + \sum_{t=2}^{10} \tau_t T_t + \varepsilon_j \tag{B2}$$

$\overline{x}_j$ average contribution in group j
UNA_j, MAJ_j dummies for team treatments (0 for IND, 1 for UNA/MAJ)
$\overline{x}_{j,I}$ average contribution by members of group j in Part I (individual 1-shot game)
T_t period dummies
ε_j composite error term, clustered at the group level
t periods 1 to 10
j group index

[17] After being assigned to a certain treatment, subjects remained in that treatment for the remainder of the experiment.

Appendix C

Table C.1.

Table C.1
Regression results – First Median Proposals versus Final Decisions.

	Part III investment in punishment	
	MAJ (1)	UNA (2)
First Median Proposals	1.034*** (0.034)	0.941*** (0.039)
Intercept	0.023 (0.056)	0.162 + (0.218)
Period dummies	yes	yes
Observations	1200	1200
Periods considered	11–20	11–20
R^2	0.842	0.809

$+p < .10$, $* p < .05$, $** p < .01$, $*** p < .001$,
OLS Model with clustered standard errors in parentheses.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.socec.2017.11.001.

References

Ambrus, A., Greiner, B., 2012. Imperfect public monitoring with costly punishment – an experimental study. Am. Econ. Rev. 102 (7), 3317–3332.
Andreoni, J., 1995. Cooperation in public-goods experiments: kindness or confusion? American Economic Review 85 (4), 891–904.
Bond, P., Eraslan, H., 2010. Strategic voting over strategic proposals. Rev. Econ. Stud. 77 (2), 459–490.
Bornstein, G., Kugler, T., Ziegelmeyer, A., 2004. Individual and group decisions in the centipede game: are groups more 'rational' players? J. Exp. Social Psychol. 40 (5), 599–605.
Bornstein, G., Yaniv, I., 1998. Individual and group behavior in the ultimatum game: are groups more 'rational' players? Exp. Econ. 1 (1), 101–108.
Cardella, E., Chiu, R., 2012. Stackelberg in the lab: the effect of group decision making and 'cooling-off' periods. J. Econ. Psychol. 33 (6), 1070–1083.
Cason, T.N., Mui, V.-L., 1997. A laboratory study in group polarisation in the team dictator game. Econ. J. 107 (444), 1465–1483.
Charness, G., Sutter, M., 2012. Groups make better self-interested decisions. J. Econ. Perspect. 26 (3), 157–176.
Cooper, D.J., Kagel, J.H., 2005. Are two heads better than one? Team versus individual play in signaling games. Am. Econ. Rev. 95 (3), 477–509.
Cox, J.C., Hayne, S.C., 2006. Barking up the right tree: are small groups rational agents? Exp. Econ. 2006 (9), 209–222.
Feddersen, T., Pesendorfer, W., 1998. Convicting the innocent: the inferiority of unanimous jury verdicts under strategic voting. Am. Polit. Sci. Rev. 92 (1), 22–35.
Feddersen, T., Pesendorfer, W., 1999. Elections, information aggregation, and strategic voting. Proc. Natl Acad. Sci. 96 (19), 10572–10574.
Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. Am. Econ. Rev. 90 (4), 980–994.
Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. Nature 415 (10 January 2002), 137–140.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.
Fischbacher, U., Gächter, S., 2010. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. Am. Econ. Rev. 100 (1), 541–556.
Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Econ. Lett. 71, 397–404.
Gächter, S., Renner, E., Sefton, M., 2008. The long-run benefits of punishment. Science 322, 1510.
Gillet, J., Schram, A., Sonnemans, J., 2009. The tragedy of the commons revisited: the importance of group decision-making. J. Public Econ. 93, 785–797.
Goeree, J.K., Holt, C.A., Laury, S.K., 2002. Private costs and public benefits: unraveling the effects of altruism and noisy behavior. J. Public Econ. 83, 255–276.
Goeree, J.K., Yariv, L., 2011. An experimental study of collective deliberation. Econometrica 79 (3), 893–921.
Greiner, B., 2003. An online recruitment system for economic experiments. In: Kremer, K., Macho, Volker (Eds.), Forschung Und Wissenschaftliches Rechnen. Ges. für Wiss. Datenverarbeitung, Göttingen, GWDG Bericht 63, pp. 79–93.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125.
Guarnaschelli, S., McKelvey, R.D., Palfrey, T.R., 2000. An experimental study of jury decision rules. Am. Polit. Sci. Rev. 94 (2), 407–423.
Herrmann, B., Thöni, C., Gächter, S., 2008. Antisocial punishment across societies. Science 319, 1362–1367.
Houser, D., Kurzban, R., 2002. Revisiting kindness and confusion in public goods experiments. Am. Econ. Rev. 92, 1062–1069.
Huber, S., Model, J., Städter, S., 2016. The Behavior of Teams and Individuals in Public Goods Games with Ostracism. University of Regensburg.
Kagel, J.H., McGee, P., 2016. Team versus individual play in finitely repeated prisoner dilemma games. Am. Econ. J. 8 (2), 253–276.
Kocher, M.G., Sutter, M., 2005. The decision maker matters: individual versus group behaviour in experimental beauty-contest games. Econ. J. 115, 200–223.
Kocher, M.G., Sutter, M., 2007. Individual versus group behavior and the role of the decision making procedure in gift-exchange experiments. Empirica 34, 63–88.
Kugler, T., Bornstein, G., Kocher, M.G., Sutter, M., 2007. Trust between individuals and groups: groups are less trusting than individuals but just as trustworthy. J. Econ. Psychol. 28 (6), 646–657.
Kugler, T., Kausel, E.E., Kocher, M.G., 2012. Are groups more rational than individuals? A review of interactive decision making in groups. WIREs Cognitive Sci. 3 (4), 471–482.
Ledyard, J.O., 1997. Public goods: a survey of experimental research. In: Kagel, J.H., Roth, A.E. (Eds.), The Handbook of Experimental Research. Princeton University Press.
Luhan, W.J., Kocher, M.G., Sutter, M., 2009. Group polarization in the team dictator game reconsidered. Exp. Econ. 12 (1), 26–41.
Müller, W., Tan, F., 2013. Who acts more like a game theorist? Group and individual play in a sequential market game and the effect of the time horizon. Games Econ. Behav. 82, 658–674.
Palfrey, T.R., Prisbrey, J.E., 1997. Anomalous behavior in public good experiments: how much and why? Am. Econ. Rev. 87, 829–846.
Robert, C., Carnevale, P.J., 1997. Group choice in ultimatum bargaining. Org. Behav. Hum. Decis. Process. 72 (2), 256–279.
Sutter, M., Kocher, M.G., Strauss, S., 2009. Individuals and teams in auctions. Oxf. Econ. Pap. 61 (2), 380–394.
Wildschut, T., Pinter, B., Vevea, J.L., Insko, C.A., Schopler, J., 2003. Beyond the group mind: a quantitative review of the interindividual-intergroup discontinuity effect. Psychol. Bull. 129 (5), 698–722.

Heike Auerswald is a member of the Center of Public and International Economics (CEPIE) and researcher at the chair of Public Economics at Technische Universität Dresden. Her research interests include public economics, environmental and resource economics, regional development and experimental economics. She obtained her Ph.D. in economics at the Technische Universität Dresden in 2014.

Carsten Schmidt is a freelance lecturer. His research focuses on experimental economics, sports economics and public economics. He obtained his Ph.D. at Humboldt University Berlin in 2001.

Marcel Thum is Professor of Economics at Technische Universität Dresden and Director of the ifo Institute of Economic Research in Dresden. His research interests include political economy and public economics. He obtained his Ph.D. at the University of Munich in 1995.

Gaute Torsvik is Associate Professor at the Department of Economics at the University of Oslo. His academic interests include labor economics, public economics and development economics. He obtained his Ph.D. in economics at the Norwegian School of Economics and Business (NHH Bergen) in 1992.